Toward Question-Asking Machines: The Logic of Questions and the Inquiry Calculus

Kevin H. Knuth*
Computational Sciences Division
NASA Ames Research Center
Moffett Field, CA 94035-1000

Abstract

For over a century, the study of logic has focused on the algebra of logical statements. This work, first performed by George Boole, has led to the development of modern computers, and was shown by Richard T. Cox to be the foundation of Bayesian inference. Meanwhile the logic of questions has been much neglected. For our computing machines to be truly intelligent, they need to be able to ask *relevant questions*. In this paper I will show how the Boolean lattice of logical statements gives rise to the free distributive lattice of questions thus defining their algebra. Furthermore, there exists a quantity analogous to probability, called *relevance*, which quantifies the degree to which one question answers another. I will show that relevance is not only a natural generalization of information theory, but also forms its foundation.

1 INTRODUCTION

Intelligent machines need to actively acquire information, and the act of asking questions is central to this capability. Question-asking comes in many forms ranging from the simplest where an instrument continuously monitors data from a sensor, to the more complex where a rover must decide which instrument to deploy or measurement to take, and even the more human-like where a robot must verbally request information from an astronaut during in an in-orbit construction task.

Intelligence is not just about providing the correct solution to a problem. When vital information is lacking, intelligence is required to formulate relevant questions. For over 150 years mathematicians have studied the logic of statements (Boole, 1854); whereas the mathematics of questions has been almost entirely neglected. In this paper, I will describe recent work performed in understanding the *algebra of questions*, and its associated calculus, the *inquiry calculus*.

Much of the material presented in this paper relies on the mathematics of partially-ordered sets and lattices. For this reason, I have included a *short appendix* to which the reader can refer for some of the mathematical background. Section §2 briefly discusses questions and the motivation for this work. Section §3 introduces the formal definition of a question. The lattice of questions and its associated algebra is described in Section §4. The question algebra is extended to the inquiry calculus in Section §5 by introducing a bi-valuation called *relevance*, which quantifies the degree to which one question answers another. In section §6, I show that the inquiry calculus is not only a *natural* generalization of information theory, but also forms its foundation. Section §7 summarizes the results, discusses how information theory has been used for some time to address question-asking, and describes how this more general methodology and deeper understanding will facilitate this process.

2 QUESTIONS

Each and every one of us asks questions, and has done so since being able to construct simple sentences. Questions are an essential mechanism by which we obtain information, however as we all know, some questions are better than others. Questions are not always verbal requests, but are often asked in the form of physical manipulations or experiments: '*What happens when I let go of my cup of milk?*' or '*Will my mother make that face again if I drop it a second time?*' Questions may also be more fundamental, such as the saccade you make when you detect motion in your peripheral visual field. Or perhaps the issue is more effectively resolved by turning your head so as to deploy

*http://ic.arc.nasa.gov/people/knuth/

both your visual and auditory sensory modalities. Regardless of their form, "questions are requests for information" (Caticha, 2004).

Many questions simply cannot be asked: there may be no one who will know the answer, no immediate way to ask it, you may not be allowed for a variety of reasons, or the question may be too expensive with respect to some cost criteria. In most situations, these questions cannot now be asked directly: '*Is there life in Europa's ocean?*', '*How fast does the SR71 Blackbird fly?*', '*How would radiation exposure on a Mars mission affect an astronaut's health?*', or '*What is the neutrino flux emitted from Alpha Centauri?*' In these cases, one must resort to asking other questions that do not directly request the information sought, yet still have relevance to the unresolved issue. This sets up the iterative process of inquiry and inference, which is essential to the process of learning—be it active learning by a machine, learning performed by a child, or the act of doing science by the scientific community.

Choosing relevant questions is a difficult task that requires intelligence. Anyone who has tried to perform a construction task with the assistance of a small child will appreciate this fact. Constantly being asked '*Do you need a hammer?*' by even the most enthusiastic helper can be a great annoyance when you are struggling to drill a hole. This is precisely the situation we will need to avoid when robots are used to assist us in difficult and dangerous construction tasks. Relevant questions asked by an intelligent assistant will be invaluable to minimizing risks and maximizing productivity in human-robot interactions. However, despite being an important activity on which we intelligent beings constantly rely, the mathematics of quantifying the relevance of a question to an outstanding issue has been surprisingly neglected.

3 DEFINING QUESTIONS

One of the most interesting facts about questions is that even though we don't know the answer, the question is essentially useless if we have absolutely no idea of what the answer could be. That is, when questions are asked intelligently, we already have a notion of the set of possible answers that the resolution may take. Richard T. Cox in his last paper captured this idea when he defined a question as the set of all logical statements that answer it (Cox, 1979).

The utility of such a definition becomes apparent when one considers the set of all possible answers to be a hypothesis space. The act of answering a question is equivalent to retrieving information, which will be used to further refine the probability density function over the hypothesis space, thereby reducing uncertainty. This can be formalized to a greater degree, and to our advantage, by realizing that a set of logical statements can be partially-ordered by the binary ordering relation '*implies*'. This set of logical statements along with its binary ordering relation $\rightarrow$, generically written in order-theoretic notation as $\leq$, forms a partially-ordered set, which can be shown to be a Boolean lattice (Birkhoff, 1967; Davey & Priestley, 2002). As a concrete example, consider a human-robotic cooperative construction task involving a robot named Bender and a human named Fry.[1] Bender has become aware that Fry will be in need of a tool, but must decide which tool Fry will prefer:

$$\begin{aligned} d &= \text{'\textit{Fry needs a drill!}'} \\ w &= \text{'\textit{Fry needs a wrench!}'} \\ h &= \text{'\textit{Fry needs a hammer!}'} \end{aligned}$$

These three atomic statements comprise the three mutually exclusive possibilities in Bender's hypothesis space. The Boolean lattice $\mathcal{A}$ (Figure 1), which I will interchangeably call the *statement lattice* or the *assertion lattice*, is the powerset of these three statements, formed by considering all possible logical disjunctions, ordered by the binary ordering relation '*implies*', $\rightarrow$. In an ideal situation, Bender's situational awareness would provide sufficient information to allow him to infer the tool Fry most probably needs. However, in reality, this will not always be the case, and Bender may need more information to adequately resolve the inference. The human way to accomplish this is to simply ask Fry for more information. Clearly, the most relevant question Bender can ask will depend both on the probabilities of the various hypotheses in this space, and on the specific issue Bender desires to resolve.

We now introduce a more formal definition of a question, which will allow us to generate a lattice of questions from a lattice of logical statements representing the hypothesis space. We first define a down-set (Davey & Priestley, 2002).

Definition 1 (Down-set) *A down-set is a subset J of an ordered set L, written $J = \downarrow L$, where if $a \in J$, $x \in L$, $x \leq a$ then $x \in J$.*

Keep in mind that $\leq$ represents the ordering relation for the ordered set—in this case $\leq$ is equivalent to $\rightarrow$ for the lattice $\mathcal{A}$. A formalized version of Cox's definition of a question follows (Knuth, 2003a, 2004b,c).

Definition 2 (Question) *A question Q is defined as a down-set of logical statements $Q = \downarrow\{a_1, a_2, \ldots, a_n\}$. The question lattice Ω is the set of down-sets of the*

[1] Bender and Fry are characters on the animated television series Futurama created by Matt Groening.

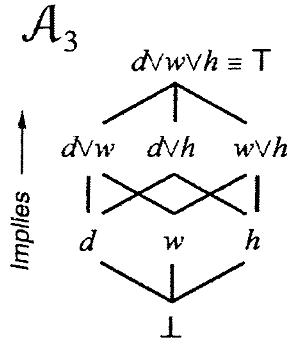

Figure 1: $\mathcal{A}_3$ is the Boolean lattice formed from three mutually exclusive assertions ordered by the relation '*implies*'. The bottom element $\perp$ is the absurdity, which is always false, and the top element $\top$ is the truism, which is always true.

assertion lattice $\mathcal{A}$ ordered by the usual set-inclusion $\subseteq$, so that $\Omega = \mathcal{O}(\mathcal{A})$.

This defines a question in terms of the set of statements that answer it, which includes all the statements that imply those statements. Note that I am using lowercase letters for assertions (logical statements), uppercase letters for questions (or sets), and script letters for ordered sets (lattices). The question lattice Ω generated from the Bender's Boolean assertion lattice $\mathcal{A}$ is shown in Figure 2 with the following notation:

$$H = \downarrow h = \{h, \perp\}$$
$$WH = \downarrow w \vee h = \{w \vee h, w, h, \perp\}$$
$$DWH = \downarrow d \vee w \vee h = \{d \vee w \vee h, \ldots\}$$

This lattice shows all the possible questions that one can ask concerning the hypothesis space $\mathcal{A}$. For example, the question $H \cup DW$ is the set union of the questions H and DW. $H \cup DW$ represents the question '*Do or do you not need a hammer?*', since this question can be answered by the statements $\{d \vee w, d, w, h, \perp\}$, where $d \vee w$ is equivalent to '*Fry does not need a hammer!*', since $\sim h = d \vee w$. [2] Note that not all of the questions in Ω have English language equivalents.

4 THE QUESTION ALGEBRA

The ordered set Ω is comprised of sets ordered by the usual set inclusion $\subseteq$. This ordering relation naturally implements the notion of *answering*. If a question A is defined by a set that is a subset of the answers to a second question B, so that $A \subseteq B$, then answering the question A will also answer the question B. Thus

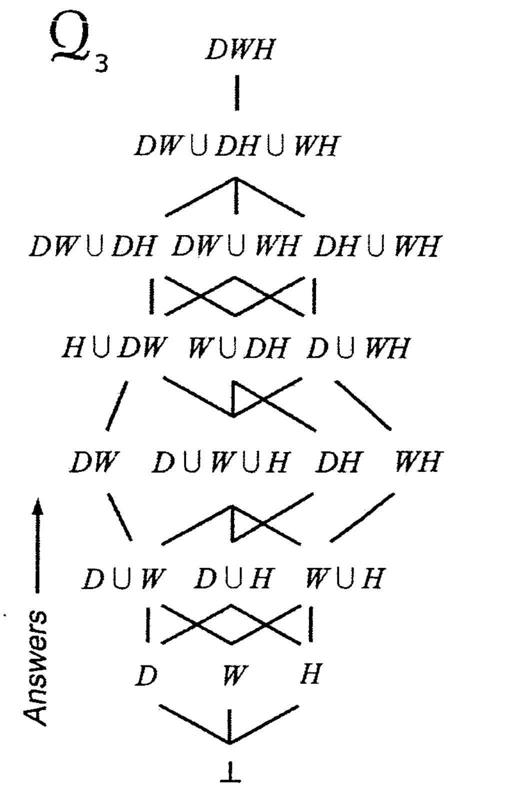

Figure 2: Ω_3 is the free distributive lattice formed from the assertion lattice $\mathcal{A}_3$. Questions are ordered by set-inclusion which implements the relation '*answers*'.

question A answers question B if and only if $A \subseteq B$. This allows us to read $A \subseteq B$ as '*A answers B*', and recognize that questions lower in the lattice (Figure 2) answer questions higher in the lattice.

The fact that the ordered set Ω is comprised of sets ordered by $\subseteq$ implies that it is a *distributive lattice* (Knuth, 2003a,b, 2004a,b,c). This means that Ω possesses two binary algebraic operations, the join $\vee$ and meet $\wedge$, which are identified with set union $\cup$ and set intersection $\cap$, respectively (Knuth, 2003a). Just as the join and meet on the assertion lattice $\mathcal{A}$ can be identified with the logical disjunction $\vee$ (OR) and the logical conjunction $\wedge$ (AND), the join and meet on the question lattice can also be viewed as a disjunction and a conjunction of questions, respectively. These operations allow us to algebraically manipulate questions as easily as we currently manipulate logical statements.

However, the similarities to the more specific Boolean algebra end there. Distributive algebras, in general, do not possess the Boolean unary operation of negation. In the case of Ω it is easy to see why, since the complement of a down-set is not a down-set. Thus, in general, questions do not possess complements.

[2]Note also that $\perp$ is the absurd answer, which answers all questions since it implies everything (see Figure 1).

Table 1: The Question Algebra

ORDERING	
Answers	$\leq \; \equiv \; \subseteq$
Reflexivity	For all A, $A \leq A$
Antisymmetry	If $A \leq B$ and $B \leq A$ then $A = B$
Transitivity	If $A \leq B$ and $B \leq C$ then $A \leq C$
OPERATIONS	
Disjunction	$\vee \; \equiv \; \cup$
Conjunction	$\wedge \; \equiv \; \cap$
Idempotency	$A \vee A = A$
	$A \wedge A = A$
Commutativity	$A \vee B = B \vee A$
	$A \wedge B = B \wedge A$
Associativity	$A \vee (B \vee C) = (A \vee B) \vee C$
	$A \wedge (B \wedge C) = (A \wedge B) \wedge C$
Absorption	$A \vee (A \wedge B) = A \wedge (A \vee B) = A$
Distributivity	$A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$
	$A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C)$
CONSISTENCY	
$A \leq B \Leftrightarrow A \wedge B = A \Leftrightarrow A \vee B = B$	

The join-irreducible elements of the question lattice $\mathcal{J}(\Omega)$ are the questions that cannot be written as the join (set union) of two other questions. These questions are called the *ideal questions* (Knuth 2003a), denoted $\mathcal{I} = \mathcal{J}(\Omega)$, reflecting the fact that they are the *ideals* (Birkhoff, 1967; Davey & Priestley, 2002) of the lattice Ω. While ideal questions are not very interesting questions to ask, they are useful mathematical constructs. The ideal questions form a lattice isomorphic to the original assertion lattice $\mathcal{A}$. Thus we have the correspondence where $\Omega = \mathcal{O}(\mathcal{A})$ and $\mathcal{A} \sim \mathcal{J}(\Omega)$. The lattices $\mathcal{A}$ and Ω are said to be *dual* in the sense of Birkhoff's Representation Theorem (Knuth 2004a). Furthermore, the operation $\mathcal{O}$ takes lattice sums to lattice products, whereas $\mathcal{J}$ takes lattice products to lattice sums. These maps are the order-theoretic analogues of the exponential and the logarithm. This will have important consequences when we generalize the question algebra to the inquiry calculus.

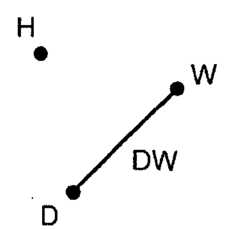

Figure 3: The hypergraph associated with the question $H \cup DW$ = '*Do you or do you not need a hammer?*'

There are other important types of questions. The first definition is originated from Cox (1979).

Definition 3 (Real Question) *A real question is a question $Q \in \Omega$, which can always be answered by a true statement. The real sublattice is denoted by $\mathcal{R}$.*

It is straightforward to show that a real question entertains each of the mutually exclusive atomic statements of $\mathcal{A}$ as acceptable answers (Knuth 2003a). This leads to the following proposition, which I will leave for the reader to prove.

Proposition 1 (The Least Real Question) *For all $Q \in \Omega$, $Q \in \mathcal{R}$ iff $Q \geq \bigvee \downarrow\mathcal{J}(\mathcal{A})$. The question $C = \bigvee \downarrow\mathcal{J}(\mathcal{A}) = \min\mathcal{R}$ is the least real question.*

Thus the least element in the real sublattice $\mathcal{R}$ is the question formed from the join of the downsets of the mutually exclusive atomic statements of $\mathcal{A}$. In our example, the least real question is $D \cup W \cup H$. This question is unique in that it answers all real questions in Ω.

Definition 4 (Central Issue) *The central issue is the least element in the real sublattice $\mathcal{R}$ of the question lattice Ω, denoted $\min\mathcal{R}$. Answering the central issue resolves all the real questions in the lattice.*

Last, a *partition question* is a real question that neatly partitions its set of answers. Specifically,

Definition 5 (Partition Question) *A partition question is a real question $P \in \mathcal{R}$ formed from the join of a set of ideal questions $P = \bigvee_{i=1}^{n} X_i$ where $\forall X_j, X_k \in \mathcal{J}(\Omega)$, $X_j \wedge X_k = \perp$ when $j \neq k$.*

There are five partition questions in our example: DWH, $H \cup DW$, $W \cup DH$, $D \cup WH$, and $D \cup W \cup H$. Together these questions form a lattice $\mathcal{P}$ isomorphic to the partition lattice $\mathbf{\Pi}_3$. Note that the central issue is the partition question with the maximal number of partitions. For this reason, it is the least ambiguous question.

The question lattice Ω generated from the Boolean lattice $\mathcal{A}$ is known as the *free distributive lattice* (Knuth, 2003a). As such, it is isomorphic to the lattice of simplicial complexes in geometry (Klain & Rota, 1997), as well as the lattice of hypergraphs (Knuth, 2004b). Thus hypergraphs are a convenient graphical means of diagramming questions. Figure 3 shows the hypergraph associated with the question $H \cup DW$ = '*Do you or do you not need a hammer?*' Such hypergraphs may play a more significant role when inquiry is united with inference in the form of Bayes Nets.

5 THE INQUIRY CALCULUS

With the question algebra well-defined, we now extend the ordering relation to a quantity that describes the *degree to which one question answers another*. This is done by defining a bi-valuation on the lattice that takes two questions and returns a real number. We call this bi-valuation the *relevance* (Knuth, 2004b).

Definition 6 (Relevance) *The degree to which a question Q resolves an outstanding issue I, for all $Q, I \in \Omega$, is called the relevance, and is written $d(I|Q)$ where*

$$d(I|Q) = \begin{cases} c & \text{if } Q \leq I \quad (Q \text{ answers } I) \\ 0 & \text{if } Q \wedge I = \bot \quad (Q \text{ and } I \text{ are exclusive}) \\ d & \text{otherwise, where } 0 < d < c. \end{cases}$$

with c being the maximal relevance.

This bi-valuation is defined so as to extend the dual of the zeta function for the lattice, which acts to quantify order-theoretic inclusion, which in this case, indicates whether the question Q answers the question I (Knuth, 2004a,b). The utility of this bi-valuation becomes apparent when one considers $I = \min \mathcal{R}$ to be the central issue, and $Q \in \mathcal{R}$ to be an arbitrary real question. The bi-valuation $d(I|Q)$ then quantifies the degree to which Q resolves the central issue I by taking a value d where $0 < d < c$.[3] This is analogous to the notation in probability theory where $p(x|y)$ describes the degree to which the statement y implies the statement x (Cox, 1946, 1961; Jaynes, 2003).

Consistency between the definition of relevance, and the lattice structure (or, equivalently, its algebra) results in three rules, which describe how relevances relate to one another (Knuth 2004a,b). Consistency with associativity gives rise to the ***Sum Rule***:

$$d(X \vee Y|Q) = d(X|Q) + d(Y|Q) - d(X \wedge Y|Q), \quad (1)$$

and its multi-question generalization

$$d(X_1 \vee X_2 \vee \cdots \vee X_n|Q) = \sum_i d(X_i|Q) - \sum_{i<j} d(X_i \wedge X_j|Q) + \sum_{i<j<k} d(X_i \wedge X_j \wedge X_k|Q) - \cdots, \quad (2)$$

which, due to the Möbius function of the distributive lattice, displays the familiar sum and difference pattern known as the *inclusion-exclusion principle* (Klain & Rota, 1997; Knuth 2004a,b).

[3]Robert Fry originally introduced the notation $b(Q|I)$, which is equivalent to my notation $d(I|Q)$. For more details, consult (Knuth, 2004b)

Consistency with distributivity of $\wedge$ over $\vee$ results in the ***Product Rule***

$$d(X \wedge Y|Q) = c\, d(X|Q) d(Y|X \wedge Q), \quad (3)$$

where the real number c is again the maximal relevance. Note that the calculus cannot simultaneously support distributivity of $\wedge$ over $\vee$ and distributivity of $\vee$ over $\wedge$, which are both allowed in a distributive lattice (Knuth 2004a,b). It may surprise some to learn that is also the case in probability theory.

Last, consistency with commutativity of $\wedge$ results in a ***Bayes' Theorem Analogue***

$$d(Y|X \wedge Q) = \frac{d(Y|Q) d(X|Y \wedge Q)}{d(X|Q)}. \quad (4)$$

The fact that these three rules are shared between the inquiry calculus and probability theory is a result of the fact that both the assertion lattice $\mathcal{A}$ and the question lattice Ω are distributive lattices, with a Boolean lattice being a special case of a distributive lattice.

Since the assertion lattice $\mathcal{A}$ and the question lattice Ω are dual to one another in the sense of Birkhoff's Representation Theorem, it is not unreasonable to expect that the values of the relevances of questions must be consistent with the probabilities of their possible answers. Given an ideal question $X = \downarrow x$ we require that

$$d(X|\top) = H(p(x|\top)), \quad (5)$$

where $d(X|\top)$ is the degree to which the question that asks everything $\top$ answers the question X, $p(x|\top)$ is the degree to which the truism (the top element of $\mathcal{A}$) implies the statement x, and H is a function to be determined. The result, which relies on partition questions (Knuth, 2004b) and an important result from Janos Aczél and colleagues (Aczél, Forte & Ng, 1974) is that the unique form of the function for a partition question $P \in \mathcal{P}$ is a linear combination of the Shannon and Hartley entropies

$$d(P|\top) = a\, H_m(p_1, p_2, \cdots, p_n) + b\, {}_oH_m(p_1, p_2, \cdots, p_n), \quad (6)$$

where $p_i \equiv p(x_i|\top)$, a, b are arbitrary non-negative constants. The Shannon entropy (Shannon & Weaver, 1949) is defined as

$$H_m(p_1, p_2, \cdots, p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i, \quad (7)$$

and the Hartley entropy (Hartley, 1928) is defined as

$${}_oH_m(p_1, p_2, \cdots, p_n) = \log_2 N(P), \quad (8)$$

where $N(P)$ is the number of non-zero arguments p_i. By setting the arbitrary constants $a = b = 1$ in (6), and using (7), (8), we get

$$H_m(p_1, p_2, \cdots, p_n) = -\sum_{i=1}^{n} p_i \log_2 \frac{p_i}{\frac{1}{n}}, \qquad (9)$$

which is the relative entropy based on a uniform measure. This result is important since it rules out the use of other entropies for the purpose of inference and inquiry. Any other entropy function will lead to an inconsistency between the bi-valuations defined on the assertion lattice $\mathcal{A}$ and the bi-valuations defined on the question lattice $\mathcal{Q}$.

6 A NATURAL GENERALIZATION OF INFORMATION THEORY

We will now show that these results not only lead naturally to information theory, but significantly generalize its scope including several generalizations already proposed in the literature. For simplicity, we will assign the arbitrary constants so that $a = 1$ and $b = 0$, and limit ourselves to the Shannon entropy. The main result of the previous section is that the degree to which the top question $\top$ answers any partition question $P \in \mathcal{P}$ is quantified by the entropy of its answers. Thus probability quantifies what we know, whereas entropy quantifies what we do not know.

However, more basic quantities also appear, and take on new fundamental importance. Since partition questions are joins of ideal questions, it is straightforward to show, using the sum rule, that the degree to which $\top$ answers an ideal question $X_i \in \mathcal{I}$ is given by the probability-weighted surprise

$$d(X_i|\top) = -p_i \log_2 p_i. \qquad (10)$$

If we look at our earlier example, we can compute the degree to which $\top$ answers the question $DW \vee WH$. This is easily done using the sum rule, which gives

$$\begin{aligned} d(DW \vee WH|\top) = d(DW|\top) + d(WH|\top) \\ - d(DW \wedge WH|\top). \quad (11) \end{aligned}$$

Clearly this quantity is related to the *mutual information* between DW and WH, which when written in standard notation would look like

$$\begin{aligned} I(DW; WH) = \\ H(DW) + H(WH) - H(DW, WH), \quad (12) \end{aligned}$$

with $d(DW \wedge WH|\top)$ being related to the *joint entropy*. Thus mutual information is related to the disjunction of two issues, whereas the joint entropy is related to the conjunction of two issues. However, in this illustration is important to note that (11) is not exactly a mutual information since neither DW nor WH are partition questions, however with a larger hypothesis space it is trivial to construct the mutual information in this way.

By considering the disjunction and conjunction of multiple issues, one can construct relevances that are higher-order mutual informations and higher-order joint entropies that exhibit the sum and difference patterns in the multi-question generalization of the sum rule. Higher-order generalizations such as these were independently suggested by several authors (McGill, 1955; Cox, 1961, 1979; Bell, 2003), however here one can see that they occur naturally as a result of the inquiry calculus.

To consider the conjunction and disjunction of questions to the right of the solidus, one must use the sum and product rules in conjunction with the Bayes' theorem analogue to move questions from one side of the solidus to the other. The following example demonstrates a typical calculation, which also includes some algebraic manipulation. Consider again Bender's central issue $T =$ '*Which tool do you need?*'. However, Bender has asked this question 10 times in the last hour, and Fry is getting quite irritated and will lose his temper if he hears that question again. To find another question, Bender computes the relevance that the question $Q_H =$ '*Do you or do you not need a hammer?*' has on the issue. This calculation results in

$$\begin{aligned} d(T|Q_H) &= d(T|Q_H \wedge \top) \qquad (13) \\ &= d(Q_H|T \wedge \top)\frac{d(Q_H|\top)}{d(T|\top)} \\ &= d(Q_H|T)\frac{d(Q_H|\top)}{d(T|\top)} \\ &= c\frac{d(Q_H|\top)}{d(T|\top)}, \end{aligned}$$

where the result is simply a ratio of two entropies. Note that this formalism relies on relevances that are conditional—like probabilities. This notion is absent in traditional information theory, and is another way in which the inquiry calculus is a natural generalization.

7 DISCUSSION

We have demonstrated that the question algebra and the inquiry calculus follow naturally from a straightforward definition of a question as the set of statements that answer it. The question algebra enables one to manipulate questions algebraically as easily as we currently manipulate logical statements, whereas the inquiry calculus allows us to quantify the degree

to which one question answers another. This methodology promises to enable us to design machines that can identify maximally relevant questions in order to actively obtain information. This work has clear implications for areas of research that rely on question-asking, such as experimental design (Lindley, 1956; Loredo, 2004), search theory (Pierce, 1979), and active learning (MacKay, 1992), each of which has taken advantage of information theory during their histories. In addition, this approach has already shown promise in several applications by Robert Fry (1995, 2002).

However, the inquiry calculus is more fundamental than information theory in the sense that it derives directly from the question algebra. The sole postulate is that the bi-valuations on the dual lattices are defined consistently. The result is that the Shannon and Hartley entropies are the only entropies that can be used for the purposes of inquiry—all other entropies will lead to inconsistencies. Entropy is related to the relevances involving the partition questions, mutual information is related to disjunctions of questions, and joint entropy is related to conjunctions of questions. Higher-order informations occur naturally when multiple disjunctions and conjunctions are considered. Last, the calculus allows for, and relies on, conditional quantities not considered in traditional information theory. The result is an algebra and a calculus that takes the guesswork out of defining information-theoretic cost functions in applications involving question-asking.

Our explorations into the realm of questions are only beginning, and it would be naïve to think that the work presented here is the entire story. Recently, Ariel Caticha presented an alternative approach to viewing a question as a probability distribution, which is in some ways simultaneously more general yet more restrictive than the approach presented here (Caticha, 2004). The result is a measure of relevance described by relative entropy. It will be interesting to see where these new investigations lead.

APPENDIX: POSETS AND LATTICES

In this section I introduce some basic concepts of order theory that are necessary to understand the spaces of logical statements and questions. Order theory captures the notion of ordering elements of a set. For a given set, one associates a *binary ordering relation* to form what is called a *partially-ordered set*, or a *poset* for short. This ordering relation, generically written $\leq$, satisfies reflexivity, antisymmetry, and transitivity. The ordering $a \leq b$ is generally read 'b includes a'. When $a \leq b$ and $a \neq b$, we write $a < b$. Furthermore, if $a < b$, but there does not exist an element x in the set such that $a < x < b$, then we write $a \prec b$, read 'b covers a', indicating that b is a direct successor to a in the hierarchy induced by the ordering relation. This concept of covering can be used to construct diagrams of a poset. If an element b includes an element a then it is drawn higher in the diagram. If b covers a then they are connected by a line.

A poset P possesses a greatest element if there exists an element $\top \in P$, called the *top*, where $x \leq \top$ for all $x \in P$. Dually, a poset may possess a least element $\bot \in P$, called the *bottom*. The elements that cover the bottom are called *atoms*.

Given two elements x and y, their *upper bound* is defined as the set of all $z \in P$ such that $x \leq z$ and $y \leq z$. If a unique *least upper bound* exists, it is called the *join*, written $x \vee y$. Dually, we can define the *lower bound* and the *greatest lower bound*, which if it exists, is called the *meet*, $x \wedge y$. Graphically the join of two elements can be found by following the lines upward until they first converge on a single element. The meet can be found dually. Elements that cannot be expressed as a join of two elements belong to a special set of elements called *join-irreducible elements*.

The dual of a poset P, written P^∂ can be formed by reversing the ordering relation, which can be visualized by flipping the poset diagram upside-down. This action exchanges joins and meets and is the reason that their relations come in pairs (see Table 1).

A *lattice* $\mathcal{L}$ is a poset where the join and meet exist for every pair of elements. We can view the lattice from a structural viewpoint as a set of objects arranged by an ordering relation $\leq$. However, we can also view the lattice from an operational viewpoint as an algebra on the space of elements with the operations $\vee$ and $\wedge$ along with any other relations induced by the ordering relation. The join and meet obey idempotency, commutativity, associativity, and the absorption property.

The act of generalizing an algebra to a calculus goes back at least as far as 1946 when Richard Cox derived probability theory by requiring consistency with the Boolean algebra of logical statements (Cox 1946, 1961). This led to the perspective where probabilities are viewed as degrees of implication and probability theory is viewed as an extension of logic (Jaynes, 2003). Ariel Caticha, working with quantum mechanical experimental setups showed that a distributive algebra can lead to sum and product rules with complex valuations (Caticha, 1998). When combined with lattice theory, this leads to a very powerful means of deriving measures from ordering relations (Knuth, 2004a).

Acknowledgements

This work was supported by the NASA IDU/IS/CICT Program and the NASA Aerospace Technology Enterprise. I am deeply indebted to Ariel Caticha, Bob Fry, Janos Aczél, and Kevin Wheeler for insightful and inspiring discussions.

References

Aczél J., Forte B. & Ng C.T. (1974). Why the Shannon and Hartley entropies are 'natural'. *Adv. Appl. Prob.*, Vol. 6, pp. 131–146.

Bell A.J. (2003). The co-information lattice. *Proceedings of the Fifth International Workshop on Independent Component Analysis and Blind Signal Separation: ICA 2003* (eds. S. Amari, A. Cichocki, S. Makino and N. Murata).

Birkhoff G.D. (1967). *Lattice Theory*, Providence:American Mathematical Society.

Boole G. (1854). *An Investigation of the Laws of Thought.* London:Macmillan.

Caticha A. (1998). Consistency, amplitudes and probabilities in quantum theory. *Phys. Rev. A*, Vol. 57, pp. 1572–1582.

Caticha A. (2004). Questions, relevance and relative entropy. In press: *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Garching, Germany, August 2004* (eds. R. Fischer, R. Preuss, U. Von Toussaint, V. Dose). AIP Conf. Proc., Melville NY:AIP.

Cox R.T. (1946). Probability, frequency, and reasonable expectation. *Am. J. Physics*, Vol. 14, pp. 1–13.

Cox R.T. (1961). *The algebra of probable inference.* Baltimore:Johns Hopkins Press.

Cox R.T. (1979). Of inference and inquiry. In *The Maximum Entropy Formalism* (eds. R. D. Levine & M. Tribus). Cambridge:MIT Press, pp. 119–167.

Davey B.A. & Priestley H.A. (2002). *Introduction to Lattices and Order.* Cambridge:Cambridge Univ. Press.

Fry R.L. (1995). Observer-participant models of neural processing. *IEEE Trans. Neural Networks*, Vol. 6, pp. 918–928.

Fry R.L. (2002). The engineering of cybernetic systems. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Baltimore MD, USA, August 2001* (ed. R. L. Fry). New York:AIP, pp. 497–528.

Hartley R.V. (1928). Transmission of information. *Bell System Tech. J.*, Vol. 7, pp. 535–563.

Jaynes E.T. (2003). *Probability theory: the logic of science.* Cambridge:Cambridge Univ. Press.

Klain D.A. & Rota G.-C. (1997). *Introduction to geometric probability.* Cambridge:Cambridge Univ. Press.

Knuth K.H. (2003a). What is a question? In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Moscow ID, USA, August 2002* (ed. C. Williams). AIP Conf. Proc. Vol. 659, Melville NY:AIP, pp. 227–242.

Knuth K.H. (2003b). Intelligent machines in the 21st century: foundations of inference and inquiry, *Phil. Trans. Roy. Soc. Lond. A*, Vol. 361, No. 1813, pp. 2859–2873.

Knuth K.H. (2004a). Deriving laws from ordering relations. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Jackson Hole WY, USA, August 2003* (ed. G. J. Erickson). AIP Conf. Proc. Vol. 707, Melville NY:AIP, pp. 204–235.

Knuth K.H. (2004b). Lattice duality: The origin of probability and entropy. In press: *Neurocomputing.*

Knuth K.H. (2004c). Measuring questions: Relevance and its relation to entropy. In press: *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Garching, Germany, August 2004* (eds. R. Fischer, R. Preuss, U. Von Toussaint, V. Dose). AIP Conf. Proc., Melville NY:AIP.

Lindley D.V. (1956). On the measure of information provided by an experiment. *Ann. Math. Statist.* Vol. 27, pp. 986–1005.

Loredo T.J. (2004). Bayesian adaptive exploration. In: *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Jackson Hole WY, USA, August 2003* (ed. G. J. Erickson). AIP Conf. Proc. Vol. 707, Melville NY:AIP, pp. 330–346.

MacKay D.J.C. (1992). Information-based objective functions for active data selection. *Neural Computation* Vol. 4 No. 4, pp. 589–603.

McGill W.J. (1955). Multivariate information transmission. *IEEE Trans Info Theory*, Vol. 4, pp. 93–111.

Pierce J.G. (1979). A new look at the relation between information theory and search theory. In *The Maximum Entropy Formalism* (eds. R. D. Levine & M. Tribus), Cambridge:MIT Press, pp. 339–402.

Shannon C.E. & Weaver W. (1949). *A mathematical theory of communication.* Chicago:Univ. of Ill. Press.